A SURVEY ON MALWARE DETECTION ON ANDROID IOT DEVICES

Ms.K. GAYATHRI

Department of computer science and engineering,

Sri Manakula Vinayagar Engineering College,

Madagadipet, Puducherry.

arungayu9396@gmail.com

Dr.E. KODHAI

Department of computer science and engineering,

Sri Manakula Vinayagar Engineering College,

Madagadipet, Puducherry.

kodhaiej@gmail.com

ABSTRACT

This survey occupies a major share in the mobile application market. Android mobiles have become an easy target for the attackers. The main reason is the user ignorance in the process of installing and usage of the apps. Android malware can be detected based on the permissions it requests from the user. Several machine learning algorithms are being used in the detection of android malware based on the list of permissions enabled for each app. This paper makes an attempt to study the performance of some of the machine learning algorithms, viz., naïve Bayes, J48, Random Forest, Multi-class classifier and Multi-layer perceptron. This application data are used for normal apps and standard malware data sets are used in the evaluation. Multi-class classifier was found to be outperforming the other algorithms in terms of classification accuracy. Naïve Bayes classifier has outperformed as far as model construction time is concerned.

Keywords: Machine learning algorithms, Malware detection, classification, permissions

INTRODUCTION

The Internet of Things (IoT) has transformed from being an interconnection of embedded computing devices to an interconnection of smart sensor devices. However, when applied to the smart city environment, it tends to open issues, such as low storage and limited processing capacity. Meanwhile, cloud computing offers fast processing and large storage capability. Therefore, we need IoT cloud integration to cope with highly demanding multimodal malware

detection services. Real-time communication and hacker exploits have always been the central idea of multimodal malware detection services. However, with IoT cloud technologies, the need for a cognitive framework that provides some real time application and high-quality multimodal malware detection at low cost increases. With artificial intelligence (AI) and deep learning techniques, introducing human-like intelligence to multimodal malware detection frameworks is timely.

The first step in an IoT application is to collect data through IoT sensors. These sensors generate raw data that need to be processed before any action can be taken. Typically, the collected data are transmitted to the base station for processing. However, often, base stations have limited processing and storage resources. In such cases, they can only carry out simple operations on the data, such as reformatting, compression/expansion, aggregation, etc. Following these operations, data are transmitted to cloud servers for further processing and decision making (i.e., to distill intelligence and, hence, impart smartness to the system). Although cloud servers have the required computational resources for signal processing and information extraction, data transmission from IoT sensors to cloud servers throws up serious design challenges, such as security concerns, insufficient energy, and limited bandwidth. To get around these obstacles, previous studies have suggested pushing data processing towards the edge of the IoT devices and implementing cryptographic techniques (i.e., encryption and hashing) on the collected data.

However, although edge-side computing enables decision making without the use of cloud resources and encryption and hashing strengthen security, the overall computational cost, and hence energy consumption, increases significantly. Therefore, simultaneously achieving smartness, security, and energy efficiency continues to be a paramount design challenge.

The detection of malware using different features information to reflect various characteristics of applications in numerous aspects. Our proposed methodology first distinguishing the malware and benign, and refines them using enhanced clustering methods. Our clustering method calculates the weights of each feature set and iterative reduced the unnecessary features that can effectively distinguish malware and benign applications even though malware have many properties similar to benign applications. Moreover, our proposed clustering method handles

multidimensional data by reducing the features using weight learning procedure. This property enables our method to fit for applying on many types of clustering problems.

In the second phase, we propose a multi-feature Naive Bayes algorithm for classification of malware. The input matrix for decision tree based Naive Bayes is the output matrix of the clustering process which includes feature-weighting matrix and clustering labels. Among many useful classification algorithms, we analyzed that the Naive Bayes is the most suitable classification approach for various types of features. Additionally, it extracts the most vital characteristics of input matrix for removing noisy objects from the data. Finally, our proposed framework integrates permission blockchain database for storing de-tracked information of malware features, which are automatically generating new blocks that can identify the new type of malware for IoT devices. The permission blockchain provides actual information in a distributed malware database to increase the runtime detection of malware more efficiently. The advantage of this method is can be applied directly to the mobile device.

TYPES OF MODELS

- **Device and location independence** enable users to access systems using a web browser regardless of their location or what device they use (e.g., PC, mobile phone). As infrastructure is off-site (typically provided by a third-party) and accessed via the Internet, users can connect from anywhere.
- **Maintenance** of cloud computing applications is easier, because they do not need to be installed on each user's computer and can be accessed from different places.
- **Multitenancy** enables sharing of resources and costs across a large pool of users thus allowing for:
  - **centralization** of infrastructure in locations with lower costs (such as real estate, electricity, etc.)
  - **peak-load capacity** increases (users need not engineer for highest possible load-levels)
  - **utilisation and efficiency** improvements for systems that are often only 10–20% utilized.

- **Performance** is monitored, and consistent and loosely coupled architectures are constructed using web services as the system interface.
- **Productivity** may be increased when multiple users can work on the same data simultaneously, rather than waiting for it to be saved and emailed. Time may be saved as information does not need to be re-entered when fields are matched, nor do users need to install application software upgrades to their computer.
- **Reliability** improves with the use of multiple redundant sites, which makes well-designed cloud computing suitable for business continuity and disaster recovery.
- **Scalability and elasticity** via dynamic ("on-demand") provisioning of resources on a fine-grained, self-service basis in near real-time (Note, the VM startup time varies by VM type, location, OS and cloud providers), without users having to engineer for peak loads. This gives the ability to scale up when the usage need increases or down if resources are not being used.
- **Security** can improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is often as good as or better than other traditional systems, in part because providers are able to devote resources to solving security issues that many customers cannot afford to tackle. However, the complexity of security is greatly increased when data is distributed over a wider area or over a greater number of devices, as well as in multi-tenant systems shared by unrelated users. In addition, user access to security audit logs may be difficult or impossible. Private cloud installations are in part motivated by users' desire to retain control over the infrastructure and avoid losing control of information security.

RELATED WORKS

K. Xu, Y. Li, R. H. Deng, and K. Chen, ``DeepRe_ner: Multi-layer android malware detection system applying deep neural networks,'' in Proc. IEEE Eur. Symp. Secur. Privacy, Apr. 2018, pp. 473_487.

It encourages the combination of public processing with service-oriented processing, giving “birth” to public Web solutions. On the one side, public processing develops user programs upon

the concepts of combined action and content discussing. However, service-oriented processing develops enterprise programs upon the concepts of support offer and demand and loose combining. Thanks to this combination public Web solutions can operate considering with whom they worked in the past and with whom they would like to work in the future. To professional public Web solutions, this document presents a four-step method that details several concerns related to the technological innovation exercise. These concerns are what connections exist between Web solutions, what public networking sites match to these connections, how to build public networking sites of Web solutions, and what public actions can Web solutions display. Encounters dealing with applying public Web solutions are, also, revealed in the document.

DISADVANTAGES

- Adversaries, such as interested service providers, can possibly make a copy of the database or eavesdrop users' queries, which will be difficult to detect and prevent in the cloud infrastructures.
- Perturbed data do not produce very accurate data mining.

ADVANTAGES

- It consists of high protection of data sharing security.
- It is reliable to monitor the data in two way communication channel.

M. Sun, X. Li, J. C. Lui, R. T. Ma, and Z. Liang, ``Monet: A user-oriented behavior-based malware variants detection system for android,'' IEEE Trans. Inf. Forensics Security, vol. 12, no. 5, pp. 1103_1112, May 2017.

The Existing theory and research on relationships among competitors focuses either on competitive or on cooperative relationships between them, and the one relationship is argued to harm or threaten the other. Little research has considered that two firms can be involved in and benefit from both cooperation and competition simultaneously and hence that both types of relationships need to be emphasized at the same time. In this article, it is argued that the most complex, but also the most advantageous relationship between competitors, is "coopetition" where two competitors both compete and cooperate with each other. Complexity is due to the fundamentally different and contradictory logics of interaction that competition and cooperation

are built on. It is of crucial importance to separate the two different parts of the relationship to manage the complexity and thereby make it possible to benefit from such a relationship. This article uses an explorative case study of two Swedish and one Finish industries where coopetition is to be found, to develop propositions about how the competitive and cooperative part of the relationship can be divided and managed. It is shown that the two parts can be separated depending on the activities degree of proximity to the customer and on the competitors' access to specific resources. It is also shown that individuals within the firm only can act in accordance with one of the two logics of interaction at a time and hence that either the two parts have to be divided between individuals within the company, or that one part needs to be controlled and regulated by an intermediate actor such as a collective association.

DISADVANTAGES

- Query forms are designed and pre-defined by developers in information management systems.
- Difficult to design a set of static query forms to satisfy various ad-hoc database queries on complex databases.

ADVANTAGES:

- We propose a dynamic query form generation approach which helps users dynamically generate query forms.
- The dynamic approach often leads to higher success rate and simpler query forms compared with a static approach.
- The ranking of form components also makes it easier for users to customize query forms.

B. MEDJAHED AND Y. ATIF, ‘‘CONTEXT-BASED MATCHING FOR WEB SERVICE COMPOSITION,’’ DISTRIB. PARALLEL DATABASES, VOL. 21, NO. 1, PP. 5-37, JAN. 2017

The Current technique about interpreting official semantics of dedication usually consider functions as axioms or constrains on top of the dedication semantics, which don't succeed to catch the significance of communications that are main to real-life business circumstances. Furthermore, existing semantic frameworks using different logics do not collect the full semantics of dedication functions and semantics of public responsibilities within the same

structure. This document produces a novel specific semantic design for public responsibilities and their functions. It suggests a sensible design based on a new reasoning increasing CTL* with responsibilities and functions to specify broker communications. We also recommend a new meaning of task and delegation functions by considering the connection between the unique and new dedication material. We confirm that the suggested design meets some qualities that are suitable when modelling broker communications in MASs and existing a Net Bill method as a operating example to explain the automated confirmation of this design. Lastly, we existing an execution and review on trial results of this method using the NuSMV and MCMAS representational design pieces.

DISADVANTAGES:

1. Existing anonymization algorithms can be used for column PSN Malefaction, e.g., Mondrian. The algorithms can be applied on the sub table containing only attributes in one column to ensure the anonymity requirement.
2. Existing data analysis (e.g., query answering) methods can be easily used on the sliced data.
3. Existing privacy measures for membership disclosure protection include differential privacy and presence.

ADVANTAGES

The load for my data warehouse has recently been moved from a series of stored process running in a SQL to running in an SSIS package to take advantage of the parallelism to the training the instructor indicated that for most tasks SQL can be quicker and more flexible.

Q. WU, A. IYENGAR, R. SUBRAMANIAN, I. ROUVELLOU, I. SILVA-LEPE, AND T. MIKALSEN, ‘‘COMBINING QUALITY OF SERVICE AND SOCIAL INFORMATION FOR RANKING SERVICES,’’ IN PROC. SERVICEWAVE WORKSHOPS CONJUNCT. 7TH ICSOC, STOCKHOLM, SWEDEN, 2017, PP. 561-575.

This survey aims to study and analyze current techniques and methods for combining quality of web service systems, to discuss future trends and propose further steps on making web services

systems context-aware. It analyzes and compares existing context-aware web service-based systems based on techniques they support, such as context information modelling, context sensing, distribution, security and privacy, and adaptation techniques. Existing systems are also examined in terms of application domains, system type, mobility support, multi-organization support and level of web services implementation. The Findings is going to Support a context-aware web service-based system is increasing. It is hard to find a truly context-aware web service-based system that is interoperable and secure, and operates on multi-organizational environments. Various issues, such as distributed context management, context-aware service modeling and engineering, context reasoning and quality of context, security and privacy issues have not been well addressed. The Originality/value – Existing surveys do not focus on context-awareness techniques for web services. This paper helps to understand the state of the art in context-aware techniques for web-services that can be employed in the future of services which is built around, amongst others, mobile devices, web services, and pervasive environments.

Disadvantages

Sentimental Emotions however comes at a significant performance cost. On smart gadgets where data, like the disgust, bad, sad etc are very limited, it is important to keep a low footprint on such solutions.

Advantages

- To achieve better performance, this project proposes to optimize the Trace oriented Algorithm which is easily recognize their users feedback with a strong validation.
- on analyzing the performance for persistent storage protection using TOA on smart gadget devices.

S. NO	AUTHOR& YEAR OF PUBLICATION	TITLE	TECHNIQUES USED	DATASET USED	PARAMETERS USED	EFFICIENCY	LIMITATIONS
1	K. Xu, Y. Li, R. H. Deng, and K. Chen 2018	Multi-layer android malware detection system	Dynamic Routing algorithm with Direction	CASIA-V4 Lamp dataset	Learning Rate (LR), Accuracy, Equal Error Rate (EER)	The accuracy of Inception V3_6bloc	The results show that the accuracy of the

		applying deep neural networks	and Length [1]			ks+DRDL network reaches 92.34%, with a decrease of only 7.33%	networks with capsule architecture decreases by approximately 15%
2	M. Sun, X. Li, J. C. Lui, R. T. Ma, and Z. Liang 2017	Monet: A user-oriented behavior-based malware variants detection system for android	CNN architecture and selective enhancement algorithm [2]	CASIA-IrisV2, CASIA-IrisV4	Receiver Operating Characteristics (ROC), True Acceptance Rate(TAR), False Acceptance Rate(FAR)	TAR performance at 10% FAR is 92.49% in the baseline cross sensor scenario, growing to 97.12% and to 98.09% in CNN	High complexity
3	B. Medjahed and Y. Atif 2017	context-based matching for web service composition	Network architecture used in CNN [3]	ND-CrossSensor-2013 and CASIA-Iris-Thousand.	Recognition Accuracy	DenseNet achieves the highest peak recognition accuracy of 98.7% at layer 6 on the LG2200 dataset and 98.8% at layer 5	Needs compression to reduce the network size, high computational complexity
4	Q. WU, A. IYENGAR, R. SUBRAMANIAN, I. ROUVELLOU, I. SILVA-LEPE, and T. Mikalsen 2017	combining quality of service and social information for ranking services	Dynamic routing algorithm [4]	MNIST	Test Error Rate(TER), Accuracy	An under-trained CapsNet achieved 99.23% accuracy on the expanded MNIST test set	Accuracy is less than DRDL

5	Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger 2017	Densely connected convolutional networks	Composite function, pooling layers [5]	CIFAR-10,CIFAR-100, CVHN and ImageNet	Test Error Rate(TER), Accuracy	On C10, reduction in error about 29% from 7.33% to 5.19%. On C100, reduction in error about 30% from 28.20% to 19.64%.	It consists of large number of layers

Research issues and opportunities

In this context, we investigated the similar approaches that the machine learning based methods to the general classification-based methods, various kinds of the Android malware detection methods were studied. The significance of the proposed framework (which is based on machine learning and blockchain) over previously designed methods, which further paves a way for its application for android malware detection in IoT devices. More precisely, the performance of this framework overpass other models in term of runtime malware detection. In addition to this, when a certain new type of malware or new feature is added the accuracy is increased further. Therefore, our proposed method can handle the multi-features by combing the advantages of machine learning and blockchain that effectively detect the malware for Android IoT devices.

Conclusion

Furthermore, the experimental results show the involved techniques can achieve higher accuracy for malware detection with a low number of false-negative and false-positive rates. Besides the significance of proposed approach towards malware detection, the limitations associated with this approach are its inability to handle some obfuscation techniques and the feature hiding techniques when decompile the apk using Dex2jar. In future, we plan to develop a framework based on the deep neural network to combine static and dynamic analysis for malware detection using blockchain. Finally, our proposed framework can be applied directly to the Android-based mobile and IoT device to achieve more security and privacy.

References

[1] J. S. Park, T. Y. Youn, H. B. Kim, K. H. Rhee, and S. U. Shin, ``Smart contract-based review system for an IoT data marketplace," Sensors, vol. 18, no. 10, p. 3577, 2018.

[2] B. L. R. Stojkoska and K. V. Trivodaliev, ``A review of Internet of Things for smart home: Challenges and solutions," J. Cleaner Prod., vol. 140, no. 3, pp. 14541464, 2017.

[3] M. Damshenas, A. Dehghantanha, K.-K. R. Choo, and R. Mahmud, ``M0Droid: An android behavioral-based malware detection model," J. Inf. Privacy Secur., vol. 11, no. 3, pp. 141157, Sep. 2015.

[4] J. Walls and K. K. R. Choo, ``A review of free cloud-based anti-malware apps for android," in Proc. 14th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun., vol. 1, Aug. 2015, pp. 10531058.

[5] H. Chen et al., ``Malware collusion attack against SVM: Issues and countermeasures," Appl. Sci., vol. 8, no. 10, p. 1718, Sep. 2018.

[6] I. Dogru and K. Ömer, ``Web-based android malicious software detection and classication system," Appl. Sci., vol. 8, no. 9, p. 1622, Sep. 2018.

[7] Strategy Analytics: Android Captures Record 88 Percent Share of Global Smartphone Shipments in Q3 2016, Businesswire, San Francisco, CA, USA, 2016.

[8] A. Demontis et al., ``Yes, machine learning can be more secure! A case study on android malware detection," IEEE Trans. Dependable Secure Comput., to be published.

[9] S. Y. Yerima, S. Sezer, and I. Muttik, ``Android malware detection using parallel machine learning classiers," in Proc. 8th Int. Conf. Next Gener. Mobile Apps, Services Technol., Sep. 2014, pp. 3742.

[10] W. Enck et al., ``TaintDroid: An information-row tracking system for realtime privacy monitoring on smartphones," Commun. ACM, vol. 57, no. 3, pp. 99106, 2014.